



A LITERATURE REVIEW ON LOAD BALANCING MECHANISMS IN CLOUD COMPUTING

D. Suresh Kumar,
Research Scholar

School of Computer Science, Engineering and Applications
Bharathidasan University
Tiruchirappalli, Tamil Nadu, India

Dr. E. George Dharma Prakash Raj,
Asst. Professor

School of Computer Science, Engineering and Applications
Bharathidasan University
Tiruchirappalli, Tamil Nadu, India

Abstract: Cloud computing is an emerging technology in the present and the days to come for providing different kind of services to the end users. It focuses mainly the dynamic services using very large scalable and virtualized resources over the Internet. The allocation of the group of resources may start a problem of availability of these resources and distributing the workload of all the VMs among themselves called as load balancing. Load balancing is a challenging phenomenon in cloud computing. It ensures that no single node is overloaded by distributing the workload among the nodes properly. This is done to improve the results of the system. Due to novelty of cloud computing field, there is no many standard load balancing algorithm used in cloud environment. Hence efficient utilization of resources must be important and for that load balancing plays a major role to get maximum benefit from the resources. Having understood the vital role of load balancing, this paper covers various load balancing algorithms related to cloud computing and shows a comparative study on load balancing algorithms.

Keywords: Cloud Computing, Load Balancing, Virtual Machine

I. INTRODUCTION

Cloud computing is a on-demand pay technology where the payment is made as per the utilization of the service. Cloud Computing has been one of the current tools embraced by both industry and academic. It provides a robustness and effectiveness to store and retrieve the data. The purpose is to achieve high resource utilization and reduction in response time while ensuring the maximum utilization of resources. It transmits all requirements through the internet dynamically when user requests, such as operating system, network, storage, software, hardware and resources. The services are mainly divided into to three types namely SaaS, PaaS and IaaS. Further, In cloud computing Everything as a Service (XaaS) takes this technology to next generation. The Cloud computing is separated into three groups such as Public, Private and Hybrid cloud. Public Cloud: A cloud is called a "public cloud" when the services are open for public use. Private Cloud: This cloud setup purposes solely for a single enterprises, whether handled internally or by a third-party, and hosted either internally or externally. Hybrid Cloud: It is a mix of both public as well as private cloud setup. [1]

This paper has been orchestrated as follows. Section I gives introduction on cloud computing. Section II states the load balancing mechanism in cloud computing. Section III illustrates the literature study on relevant load balancing algorithms. Section IV describes performance analysis of load balancing algorithms. Section V gives the conclusion.

II. CLOUD COMPUTING LOAD BALANCER

Load balancer is a component in cloud OS that delivers high resource time and effective resource consumption by allocating the total load among the various cloud nodes, side

by side it solves the problem of exploitation of virtual machines. Load balancing provides solution to the problem of overburdening and focuses on higher throughput, boosting resource utilization and diminishing the response time. It is the prerequisite for maximizing the cloud performance and utilizing the resources powerfully. The utilization of clouds has been enhanced by a resource allocation method which has pre-emptible task execution. The load balancing is an efficient and critical concept in cloud computing and it helps to utilize the resources optimally, thereby minimizing the utilization of resources. Thus load will be distributed over the nodes in cloud-based Model, so that each resource does the same amount of work at any point of time which is allocated by a load balancer. The load balancer governs the various request allocation to different servers. The load balancer uses various algorithm to regulate the server which has to handle the demand.[1] The proposes load balancing algorithms are to achieve high throughput and minimizing the response time.

III- RELATED WORKS

Round Robin VMLB Algorithm

It is a basic load balancing mechanism that keeps the fresh arriving cloudlets on the accessible virtual machines in a circular manner. The major benefit of this algorithm is easiness and easy operation. The main problems are prior knowledge of user tasks and system resources. Further, it does not utilize current state of the system. [2]

Throttled Virtual Machine Load Balancing Algorithm

In this dynamic approach a user submits request to the Data Center Controller (DCC). Then, Data Center Controller requests the VM Load Balancer to identify the suitable virtual machine that can handle the given workload easily. Throttled VM Load Balancer preserves a virtual machine list and their

status (available/busy). If a suitable VM is seen on memory space, cores or availability basis, then throttled VM Load Balancer takes the cloudlet request and allocates the cloudlet request over that virtual machine. Otherwise, client has to wait in the queue until an appropriate VM becomes accessible. Among all, it is finest method for load balancing, since it upholds the present state of all VMs in data center. But the major disadvantage is that it works properly only if all VMs in a data center have same hardware structure. [2]

ESCE VM load Balancing Algorithm

ESCE means Equally Spread Current Execution. It is also called known as Active VM Load Balancing algorithm. This algorithm is based on spread spectrum technique. It equally allocates the workload among VMs in data center. A job queue retains all the cloudlet requests that need the VM for their execution. ESCE VM Load Balancer (VMLB) also upholds a list of virtual machines. The VM Load Balancer continuously verifies the job queue and VM list. If a VM is found free, then cloudlet request will be chosen over that VM. At the same time, VMLB examine the overloaded VMs. If any virtual machine is found burdened, then VMLB move some load to an idle or an under loaded virtual machine, so as to decrease some load of overloaded VM. The main downside is high computational overhead. [2]

An enhanced priority based HTV load balancing algorithm

This algorithm accomplishes an effective and trustworthy resource distribution of the tasks on the servers in data centers. This algorithm focuses on three parameters such as server load, stats of server and time constraint of the tasks. This algorithm calculated the load and performance factor of each virtual machine and then allocates the arriving task to various virtual machines according to their completion time and stand-by time to rise the throughput and performance. [3]

Virtual Machine and Physical Machine Categorization Algorithm

In general, load balancing approach that amplify the physical resource utilization and curtail the energy consumption. To calculate the performance of the approach it is compared with the existing load balancing approach and judged against the number of migration and energy consumption. Experiment results say that this approach gives better result while it is compared with the existing load balancing methods. [4]

Composite procedure

Load Balancing is required to appropriately undertake the resources of the service contributor. Load balancer is a component of cloud fabric to dispense the workload among many virtual machines in a Server over the network to accomplish optimal resource consumption, reduction in data processing time, and decline in average response time, and eliminate the overload. This composite approach is effectively utilized to handle the load distribution techniques very well. This method makes use of the advent of Equally Spread Current Execution (ESCE) and Throttled algorithms. [5]

Distributed dynamic priority based algorithm

It is used for balancing the load on instances effectively, improves the system consistency, minimizes response time and increases the throughput. Allocating the resources on virtual machines based on priority achieves the better response time and processing time. Load balancing guarantees all occurrences in a node in the networks to do the identical amount of work at any instant of time. Priority based resource provision to improve the utilization of resources and reducing response time of cloud services. [6]

Procedure : Burstness-aware load balancing

This procedure takes two load balancing algorithms. In which, RR is used for burst and Random is used in non-burst state. In order to handle the user request, Fuzzy logic based validation is used and the request is forward to a well-balanced VM. Experimental results demonstrate that the algorithm advances the average response time and average processing time in contrast with other algorithms. [7]

Honey Behavior Load Balancing method

The high priority tasks are detached from overloaded virtual machine and they are fixed to under loaded virtual machine by bearing in mind least numbers of similar priorities to those tasks, cost effective virtual machine, and least probable completion time which also balances the loads of reliant on tasks in pre-emptive manner. The minimum expected finishing time, cost and priority at submission time of that task aids to yield minimum completion time, declines waiting time of the tasks and eventually reaches better resource utilization. [8]

Procedure: Weighted based optimized load balancing

The loads are distributed among the servers which spawns VMs by considering weight as a parameter. The results are compared with existing Round Robin and EIPR algorithms. Simulation results have affirmed that this algorithm has disseminated the work load evenly among virtual machines. [9]

New enhanced load balancing procedure

Firstly, an adaptive strategy has been developed for load balancing according to the quality of the solutions discovered by Genetic. Secondly, the improved load balancing strategy takes other constraints like fitness and the choice of the initial resource pool. These constraints offer the significant impact on the result of the algorithm. This enhanced load balancing algorithm provides the better result than the existing genetic approach. [10]

Tow –level global load balancing Framework

A framework existing to represent the global server load balancing of the Web sites in a cloud with the help of two-level load balancing architecture. This framework depicts the load balancing mechanism on different data centers managed by cloud service providers.[11] Further, the method of handling the request with or without priority under the give work load also shown as client request and load balancer.

Procedure: Dynamic load balancing

Load balancing algorithms play important role in equalizing load among data centers and in efficient use of computing resources. In this paper, performance of a dynamic load balancing algorithm has been evaluated by dividing data-centers in different zones. It has been shown that the proposed algorithm improves the computing efficiency of data-centers and minimizes the response time of user's applications. [12]

IV- PERFORMANCE ANALYSIS OF LOAD BALANCING ALGORITHMS

The forthcoming part discusses various load balancing algorithms and shows the results comparatively.

Min- min LB procedure

Min-Min is a basic, simple and fastest algorithm which is capable of giving better performance. Min-Min uses the ideal tasks at initial stage, which results in best and attaining the progress of the overall makespan. The assignment of small task first is considered as an issue. This method gives focus to minimum completion tasks rather than longer completion time. This phenomenon may lead to underutilization of resources. Min-Min displays minimum completion time for jobs which are not allocated and later depute the jobs with minimum completion time (hence min-min) to a node that is talented to manage the task. [13]

Max-min LB procedure

At first for all the existing tasks are produced to the system and minimum accomplishment time for all of them are calculated, then among these tasks the one which is having the completion time, the maximum is selected and that is assigned to the corresponding machine. If the task is a single long time consuming task then, Max-Min algorithm runs small tasks parallel along with the long time consuming task. It further considers the task with maximum completion time whereas Min-Min considers the minimum completion time at first. This algorithm performance is affected due to starvation where the tasks are having the extreme finishing time will get performed first while leaving behind the tasks having the minimum completion time. [13]

RASA procedure

Resource Aware Scheduling Algorithm takes the results of Min-Min and Max-Min algorithm alternatively in order to give improved results. The resource finishing time is measured followed by Min-Min and Max -Min algorithms are utilized alternatively. [13]

Minimum Makespan procedure

This algorithm first verifies the completion of all tasks with minimum makespan and it relates with the task and migrated. The two tasks give the same makespan time it selects the node with higher performance resources. [13]

PA-LBIMM Algorithm

PA-LBIMM priority aware load balancing improved min min algorithm separate the tasks into G1 and G2 groups. The tasks submitted by VIP user's or high priority users are considered as group G1 and tasks submitted by low priority users are considered as group G2. Tasks are planned to the resources on the basis of priority. At first, for all the tasks in G1, each task is allocated to the VIP category resource with the help Min-Min, and then each task in G2 group is given to all the resources by using Min-Min. [14]

RPA-LBIMM Algorithm

This algorithm RPA-LBIMM recovery priority aware load balancing improved min-min algorithm is also similar to the priority aware load balancing improved min-min algorithm. Here use the recovery policy which helps the cloud scheduler to reschedule the tasks if a resource fails at the time of execution to achieve the minimum makespan. According to this policy, First of all, scheduler looks for the failed resource. All the tasks that were scheduled by PA-LBIMM to execute on will be considered as a task set. [14]

TABLE 1. Cloud Computing Load balancing Mechanisms

S. No.	Load balancing alg.	Parameters Used
1	Min Min	Makespan, Resource utilization
2	MaxMin	Maekspan, Resource utilization
3	RASA	Maekspan , Resource utilization
4	Minimum Makespan	Maekspan , Resource utilization
5	PA-LBIMM	Maekspan
6	RPA-LBIMM	Maekspan

Table 1 lists out the various load balancing algorithms and the parameters used in the algorithms. All these algorithms use the parameters completion time and resources utilization in cloud environment. These algorithms distribute the load based on the number of user request and number of available resource in cloud and also considering the proper utilization of resource.

TABLE 2. Task Completion Time makespan (sec.) of the given algorithms

S. No.	Load balancing alg.	Completion time Makespan (sec.)
1	Min Min	30.35
2	MaxMin	20.03
3	RASA	11.45
4	Minimum Makespan	10.8
5	PA-LBIMM	38.37
6	RPA-LBIMM	35

Table 2 represents the completion time of the task based on the number of nodes in the simulation area. The makespan of the different algorithms are given.

TABLE 3. Utilization of resources by various algorithms

S. No.	Load balancing alg.	Resource utilization (%)
1	Min Min	43.09
2	MaxMin	79.32
3	RASA	88.72
4	Minimum Makespan	91.01
5	PA-LBIMM	--
6	RPA-LBIMM	--

Table 3 represents the resource utilization of various load balancing algorithms. The average resource utilization of the different algorithms is given.

II. V - CONCLUSION

This study deals with various load balancing algorithms. The existing algorithms are static, dynamic, composite, and prioritized. The ultimate purpose of those algorithms is to reduce the response time and maximize the resource utilization. The results of the previous algorithms are limited to give improved result. Still there is plenty of space to improve the results to extract best service from cloud service providers. This study also shows the comparative results of the existing load balancing algorithms based on the parameters such as makespan and resource utilization.

REFERENCES

- [1] Reena Panwar Bhawna Mallick, "A Comparative Study of Load Balancing Algorithms in Cloud Computing", International Journal of Computer Applications (0975 – 8887) Volume 117 – No. 24, May 2015.
- [2] Mamta Khanchi, Sanjay Tyagi, "An Efficient Algorithm for Load Balancing in Cloud Computing" International Journal of Engineering Sciences & Research Technology, June, 2016.
- [3] Divya Garg, Urvashi Saxena," Dynamic Queue Based Enhanced HTV Dynamic Load Balancing Algorithm in Cloud Computing", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 5, Issue 1, January 2016.
- [4] Sachin Soni, Praveen Yadav, "A Load Balancing Approach to Minimize the Resource Wastage in Cloud Computing", International Advanced Research Journal in Science, Engineering and Technology, Vol. 3, Issue 3, March 2016.
- [5] Navtej Singh Ghumman, Rajesh Sachdeva," an efficient approach for load balancing in cloud computing using composite techniques", International Journal of Research in Engineering and Applied Sciences, volume 6, issue 2 February, 2016.
- [6] G.Suryadevi, D.Vijayakumar, R.SabariMuthuKumar, Dr. K .G. Srinivasagan, "An Efficient Distributed Dynamic Load Balancing Algorithm for Private Cloud Environment", International Journal of Innovative Research in Science, Engineering and Technology, Volume 3, Special Issue 3, March 2014.
- [7] Sally F. Issawi, Alaa Al Halees, Alaa Al Halees, "An Efficient Adaptive Load Balancing Algorithm for Cloud Computing Under Bursty Workloads", Engineering, Technology & Applied Science Research Vol. 5, No. 3, 2015.
- [8] Khushbu Zalavadiya,Dinesh Vaghela , "Honey Bee Behavior Load Balancing of Tasks in Cloud Computing", International Journal of Computer Applications (0975 – 8887), Volume 139 – No.1, April 2016.
- [9] B. Nithya Nandhalakshmi , Mahalingam, "Efficient Load Balancing in Cloud Computing Using Weighted Throttled Algorithm", International Journal of Innovative Research in Computer and Communication Engineering ,Vol. 3, Issue 6, June 2015.
- [10] Er. Pooja Er. Vivek Thapar,"An Enhanced Virtual Machine Load Balancing Algorithm for Cloud Environment", , International Journal of Emerging Research in Management &Technology, ISSN: 2278-9359 (Volume-5, Issue-5),May 2016.
- [11] Po-Huei Liang1 and Jiann-Min Yang. "Evaluation of two-level Global Load Balancing Framework in cloud environment" International Journal of Computer Science & Information Technology (IJCSIT) Vol 7, No 2, April 2015.
- [12] Navtej Singh Ghumman, Rajesh Sachdeva," an efficient approach for load balancing in cloud computing using composite techniques", International Journal of Research in Engineering and Applied Sciences, volume 6, issue 2 February, 2016.
- [13] Er. Rajeev Mangla, Er. Harpreet Singh," Recovery and user priority based load balancing in cloud computing", International Journal of Engineering and Science and Research, February 2015.
- [14] Harish Chandra, Pradeep Semwal, Sandeep Chopra," load balancing in cloud computing using a novel minimum makespan algorithm", International Journal of Advanced Research in Computer Engineering & Technology ,Volume 5, Issue 4, April 2016.